

Examen Final - Técnicas Multivariadas

Grupo 4

2025-12-08

1 - Modelo de R. Log-Binomial y Modelo de R. Poisson Robusto

Modelo de Regresión Log-Binomial

Modelo de Regresión Poisson Robusto

2 - Modelo de Regresión Logística Multinomial

1. Contexto y Objetivo del Modelo

- **Estudio:** Factores asociados al estado nutricional de niños en edad preescolar (2-5 años) en comunidades rurales.
- **Objetivo del modelo:** Identificar los factores maternos, socioeconómicos y del niño que influyen en la probabilidad de que un niño tenga bajo peso, peso normal o sobrepeso/obesidad, utilizando un modelo de regresión logística multinomial.
- **Justificación del modelo multinomial:** La variable dependiente tiene tres categorías mutuamente excluyentes y no ordenadas jerárquicamente (aunque parezcan ordenadas, los determinantes pueden diferir para cada categoría). Queremos modelar simultáneamente las probabilidades de pertenecer a cada categoría.

Simulación de la data

```
## 'data.frame': 500 obs. of 12 variables:
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ estado_nutricional: Factor w/ 3 levels "bajo_peso","normal",...: 2 3 1 2 2 2 2 1 2 ...
## $ edad_meses : num 34 52 39 56 58 26 43 56 44 40 ...
## $ sexo : Factor w/ 2 levels "femenino","masculino": 2 2 2 2 2 2 1 2 1 2 ...
## $ peso_nacimiento : num 2.7 3.5 2.4 4.1 4.1 3.2 3.9 2.7 2.2 3.1 ...
## $ edad_madre : num 43 45 30 24 37 33 34 30 30 34 ...
## $ educacion_madre : Factor w/ 3 levels "primaria","secundaria",...: 2 2 2 1 1 1 2 2 2 2 ...
## $ imc_madre : num 32.5 29.3 30.5 23.4 29.4 25.9 29.1 31.4 33.3 27 ...
## $ ingreso_familiar : num 406 1143 579 826 384 ...
## $ zona : Factor w/ 2 levels "rural","urbana": 2 2 1 2 2 2 2 2 1 1 ...
## $ lactancia_materna : num 7 20 14 19 7 3 21 0 14 23 ...
## $ dietas_solidos : Factor w/ 2 levels "no","si": 1 1 1 1 1 1 1 1 2 2 ...

## id estado_nutricional edad_meses sexo
## Min. : 1.0 bajo_peso:150 Min. :24.00 femenino :248
## 1st Qu.:125.8 normal :290 1st Qu.:33.00 masculino:252
## Median :250.5 sobrepeso: 60 Median :41.00
## Mean :250.5 Mean :41.83
## 3rd Qu.:375.2 3rd Qu.:50.00
## Max. :500.0 Max. :60.00
## peso_nacimiento edad_madre educacion_madre imc_madre
## Min. :2.000 Min. :18.00 primaria :192 Min. :18.10
## 1st Qu.:2.600 1st Qu.:25.00 secundaria:232 1st Qu.:22.27
## Median :3.200 Median :31.00 superior : 76 Median :26.60
## Mean :3.235 Mean :31.57 Mean :26.61
## 3rd Qu.:3.800 3rd Qu.:38.00 3rd Qu.:30.90
## Max. :4.500 Max. :45.00 Max. :35.00
## ingreso_familiar zona lactancia_materna dietas_solidos
## Min. : 201.0 rural :206 Min. : 0.00 no:358
## 1st Qu.: 450.8 urbana:294 1st Qu.: 6.00 si:142
## Median : 704.0 Median :12.00
## Mean : 700.7 Mean :11.77
## 3rd Qu.: 953.8 3rd Qu.:17.00
## Max. :1199.0 Max. :24.00
```

##	id	estado_nutricional	edad_meses	sexo	peso_nacimiento	edad_madre
## 1	1	normal	34	masculino	2.7	43
## 2	2	sobrepeso	52	masculino	3.5	45
## 3	3	bajo_peso	39	masculino	2.4	30
## 4	4	normal	56	masculino	4.1	24
## 5	5	normal	58	masculino	4.1	37
## 6	6	normal	26	masculino	3.2	33

##	educacion_madre	imc_madre	ingreso_familiar	zona	lactancia_materna
## 1	secundaria	32.5	406	urbana	7
## 2	secundaria	29.3	1143	urbana	20
## 3	secundaria	30.5	579	rural	14
## 4	primaria	23.4	826	urbana	19
## 5	primaria	29.4	384	urbana	7
## 6	primaria	25.9	859	urbana	3

##	dietas_solidos
## 1	no
## 2	no
## 3	no
## 4	no
## 5	no
## 6	no

Descripción de variables

Variable Dependiente:

- **estado_nutricional:** Categoría nutricional del niño (3 niveles)
 - bajo_peso: Peso/edad < -2 desviaciones estándar (Z-score OMS)
 - normal: Peso/edad entre -2 y +1 desviaciones estándar
 - sobrepeso: Peso/edad > +1 desviaciones estándar

Variables Independientes:

1. Características del niño:

- **edad_meses:** Edad en meses (continuo, 24-60 meses)
- **sexo:** Sexo biológico (categórico: **femenino**, **masculino**)
- **peso_nacimiento:** Peso al nacer en kg (continuo, 2.0-4.5 *kg*)

2. Características maternas:

- **edad_madre:** Edad de la madre en años (continuo, 18-45 años)
- **educacion_madre:** Nivel educativo (categórico: **primaria**, **secundaria**, **superior**)
- **imc_madre:** Índice de Masa Corporal materno (continuo, 18-35 *kg/m²*)

3. Factores socioeconómicos:

- **ingreso_familiar:** Ingreso mensual familiar en USD (continuo, 200-1200)
- **zona:** Zona de residencia (categórico: **urbana**, **rural**)

4. Prácticas de alimentación:

- **lactancia_materna:** Meses de lactancia materna exclusiva (continuo, 0-24 meses)
- **dietas_solidos:** Introducción de sólidos antes de 6 meses (categórico: **si**, **no**)

2. Análisis Exploratorio

Characteristic	Overall N = 500	bajo_peso N = 150	normal N = 290	sobrepeso N = 60	p-value
edad_meses	42 (10)	43 (10)	41 (10)	41 (11)	0.3
sexo					0.7
femenino	248 (50%)	78 (52%)	142 (49%)	28 (47%)	
masculino	252 (50%)	72 (48%)	148 (51%)	32 (53%)	
peso_nacimiento	3.24 (0.71)	3.17 (0.74)	3.25 (0.69)	3.33 (0.71)	0.3
edad_madre	32 (8)	33 (7)	31 (8)	31 (8)	0.021
educacion_madre					0.072
primaria	192 (38%)	61 (41%)	106 (37%)	25 (42%)	
secundaria	232 (46%)	60 (40%)	140 (48%)	32 (53%)	
superior	76 (15%)	29 (19%)	44 (15%)	3 (5.0%)	
imc_madre	26.6 (4.9)	26.7 (4.6)	26.5 (5.0)	27.1 (5.1)	0.7
ingreso_familiar	701 (287)	637 (290)	730 (281)	719 (290)	0.006
zona					0.008
rural	206 (41%)	77 (51%)	104 (36%)	25 (42%)	
urbana	294 (59%)	73 (49%)	186 (64%)	35 (58%)	
lactancia_materna	12 (7)	12 (6)	12 (7)	11 (7)	0.8
dietas_solidos					0.6
no	358 (72%)	103 (69%)	211 (73%)	44 (73%)	
si	142 (28%)	47 (31%)	79 (27%)	16 (27%)	

¹ Mean (SD); n (%)

² Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Características generales de la muestra

- La muestra consta de 500 niños en edad preescolar (24-60 meses)
- **Distribución del estado nutricional:**
 - **58% (n=290)** tiene peso normal (categoría de referencia)
 - **30% (n=150)** presenta bajo peso
 - **12% (n=60)** tiene sobrepeso
- **Distribución equilibrada por sexo:** 50% femenino, 50% masculino

Diferencias significativas entre grupos

1. Variables con diferencias estadísticamente significativas ($p < 0.05$):

- **Edad de la madre ($p = 0.021$)**
 - Las madres de niños con **bajo peso** son significativamente mayores (33 ± 7 años)
 - Las madres de niños con **peso normal y sobrepeso** son más jóvenes (31 ± 8 años)
 - **Interpretación:** La mayor edad materna podría asociarse con prácticas alimentarias menos adecuadas o factores biológicos que predisponen al bajo peso infantil.

- **Ingreso familiar ($p = 0.006$)**
 - Familias de niños con **bajo peso** tienen menor ingreso (637 ± 290 USD)
 - Familias de niños con **peso normal** tienen mayor ingreso (730 ± 281 USD)
 - **Interpretación:** El nivel socioeconómico se asocia significativamente con el estado nutricional, siendo el bajo ingreso un factor de riesgo para el bajo peso.
- **Zona de residencia ($p = 0.008$)**
 - **51%** de niños con bajo peso viven en zona rural.
 - Solo **36%** de niños con peso normal viven en zona rural.
 - **Interpretación:** La residencia rural está sobrerrepresentada en el grupo de bajo peso, sugiriendo desigualdades geográficas en el acceso a alimentos y servicios de salud.

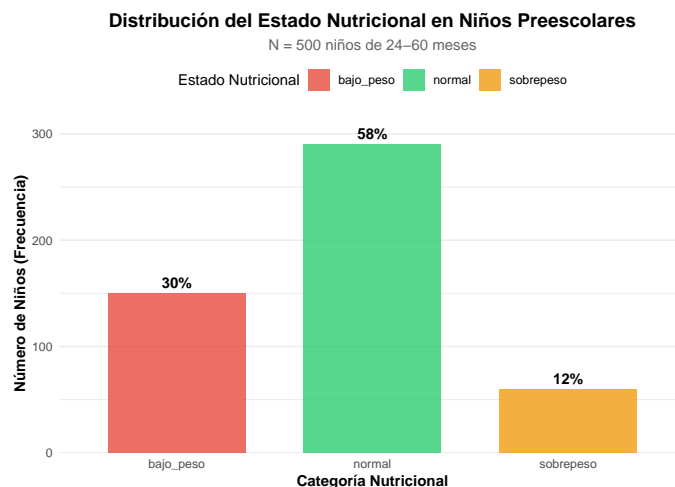
2. Variables con Diferencias Marginalmente Significativas ($p < 0.10$):

- **Educación materna ($p = 0.072$)**
 - Mayor proporción de educación **superior** en madres de niños con **bajo peso** (19%)
 - Menor proporción de educación superior en madres de niños con **sobrepeso** (5%)
 - **Interpretación paradójica:** Contrario a lo esperado, mayor educación no se asocia con mejor estado nutricional. Esto merece investigación adicional.

3. Variables sin diferencias significativas:

- **Edad del niño:** Similar en los tres grupos (~42 meses)
- **Sexo:** Distribución uniforme en todas las categorías
- **Peso al nacer:** Promedio similar (3.2-3.3 kg)
- **IMC materno:** No difiere entre grupos (~26.6 kg/m²)
- **Lactancia materna:** Similar duración (~12 meses)
- **Introducción temprana de sólidos:** Proporciones similares (~28%)

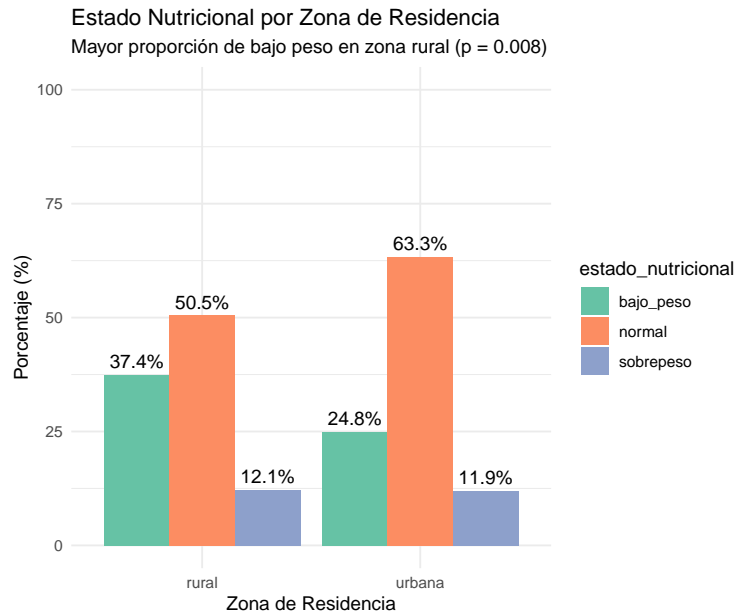
Gráfico de distribución



Interpretación:

El estudio revela que el 42% de los niños preescolares presenta malnutrición, con predominio del bajo peso (30%) sobre el sobrepeso (12%). Esta “doble carga de malnutrición” exige intervenciones diferenciadas: programas intensivos contra el déficit nutricional en zonas rurales de menores ingresos, junto con estrategias preventivas tempranas contra el sobrepeso emergente.

Relación con algunas variables clave (Zona de residencia)



Interpretación:

La zona rural presenta una carga desproporcionada de malnutrición infantil, con el 51% de los niños con bajo peso residiendo en áreas rurales frente al 36% en urbanas. Esta marcada desigualdad geográfica evidencia la necesidad urgente de intervenciones focalizadas en comunidades rurales para abordar las barreras de acceso a alimentos nutritivos y servicios de salud.

3. Modelo Multinomial

```
## # weights: 39 (24 variable)
## initial value 549.306144
## iter 10 value 485.435228
## iter 20 value 445.767698
## iter 30 value 440.447422
## final value 440.446953
## converged
```

Información técnica del ajuste

- **Algoritmo:** Red neuronal con 39 pesos
- **Variables:** 24 parámetros estimados
- **Convergencia:** EXITOSA en 30 iteraciones
- **Log-verosimilitud final:** 440.45 (mejoró desde 549.31 inicial)

Interpretación:

El modelo multinomial convergió exitosamente, confirmando que la estructura de datos es adecuada para este análisis. Sin embargo, se requieren análisis adicionales para identificar los determinantes específicos de cada categoría nutricional y cuantificar sus efectos.

Resumen del modelo

```
## Call:
## nnet::multinom(formula = estado_nutricional ~ ., data = datos)
##
## Coefficients:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal      1.112948 -0.01673057   0.1986707      0.1645671 -0.04418355
## sobrepeso    -1.392181 -0.01737771   0.3031393      0.3683165 -0.04765316
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              0.2319787              -0.1881439 -0.01081198
## sobrepeso           0.1921715              -1.5594876  0.02676942
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal      0.001292066  0.7094980      0.004338437      -0.2618619
## sobrepeso    0.001088002  0.4413381      -0.012771033      -0.2914730
##
## Std. Errors:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal      0.7692104 0.00987591   0.2129029      0.1429188 0.01363424
## sobrepeso    0.3364345 0.01395788   0.3179765      0.2008085 0.01976922
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              0.2314539              0.3049606 0.02045673
## sobrepeso           0.3277413              0.6625707 0.02728045
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal      0.0003741659 0.2127896      0.01573123      0.2310006
## sobrepeso    0.0005571530 0.3207917      0.02320365      0.3515659
##
## Residual Deviance: 880.8939
## AIC: 928.8939
```

Obtener odds ratios

```
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal      3.0433173 0.9834086   1.219780      1.178883 0.9567783
## sobrepeso    0.2485327 0.9827724   1.354103      1.445299 0.9534644
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal              1.261093              0.8284955 0.9892463
## sobrepeso           1.211878              0.2102438 1.0271309
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal      1.001293  2.032970      1.0043479      0.7696173
## sobrepeso    1.001089  1.554786      0.9873102      0.7471622
```

Evaluar significancia global (test de razón de verosimilitud)

```
## # weights: 6 (2 variable)
## initial value 549.306144
## final value 465.782614
## converged
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: estado_nutricional
##
## 1
## 2 edad_meses + sexo + peso_nacimiento + edad_madre + educacion_madre + imc_madre + ingreso_familiar + zona + lactancia_materna + dietas_solidos
##   Resid. df Resid. Dev   Test      Df LR stat.      Pr(Chi)
## 1       998   931.5652
## 2       976   880.8939 1 vs 2      22 50.67132 0.0004755304
```

Interpretación:

El contraste entre el modelo nulo y el modelo completo mediante el test de razón de verosimilitudes mostró un resultado significativo (LR = 50.67; gl = 22; p = 0.00048).

Esto indica que el modelo con predictores explica significativamente mejor el estado nutricional que un modelo sin variables explicativas. Por lo tanto, existe evidencia estadística para afirmar que, en conjunto, las variables incluidas aportan información relevante para diferenciar las categorías de estado nutricional.

4. Validación de Supuestos del Modelo

En esta sección se verifica que los datos cumplan con los requisitos teóricos para que las estimaciones sean confiables y se justifica la elección del modelo multinomial.

4.1. Evaluación de la Adecuación del Modelo: Multinomial vs. Ordinal

- **Objetivo:** Determinar si, dado el orden natural aparente de la variable dependiente, un modelo de regresión logística ordinal sería más apropiado.
- **Análisis realizado:** Prueba de Brant (Test de proporcionalidad de odds).

H_0 : Los odds son proporcionales (es decir, el modelo ordinal es adecuado).

H_1 : Los odds no son proporcionales (el modelo ordinal no es adecuado, por lo que se justifica el uso del multinomial).

```
## Call:
## MASS::polr(formula = estado_nutricional ~ edad_meses + sexo +
##   peso_nacimiento + edad_madre + educacion_madre + imc_madre +
##   ingreso_familiar + zona + lactancia_materna + dietas_solidos,
##   data = datos, Hess = TRUE)
##
## Coefficients:
##          edad_meses          sexomasculino          peso_nacimiento
##          -0.0140610388          0.2034078966          0.2017014982
##          edad_madre educacion_madresecondaria educacion_madresuperior
##          -0.0364619525          0.1547884760          -0.5404757648
##          imc_madre          ingreso_familiar          zonaurbana
##          0.0084937790          0.0009019769          0.4416759552
##          lactancia_materna          dietas_solidossi
##          -0.0049222964          -0.1895356032
##
```



```
## Intercepts:
## bajo_peso|normal normal|sobrepeso
##      -0.878221      2.124117
##
## Residual Deviance: 895.8913
## AIC: 921.8913

## -----
## Test for          X2 df probability
## -----
## Omnibus              15.16  11  0.18
## edad_meses           0.52   1  0.47
## sexomasculino         0.02   1  0.88
## peso_nacimiento       0.05   1  0.83
## edad_madre            1.78   1  0.18
## educacion_madresecundaria 0.34   1  0.56
## educacion_madresuperior 2.69   1  0.1
## imc_madre             1.48   1  0.22
## ingreso_familiar      3.72   1  0.05
## zonaurbana             5.13   1  0.02
## lactancia_materna      0.6   1  0.44
## dietas_solidossi       0.19   1  0.66
## -----
##
## H0: Parallel Regression Assumption holds
```

Decisión estadística: Dado que $p = 0.18 > 0.05$, **no se rechaza** H_0 . Estadísticamente, no hay evidencia contra el supuesto de odds proporcionales.

Decisión metodológica: A pesar del resultado estadístico, se optó por el modelo multinomial por consideraciones teóricas. La literatura sugiere que los determinantes del bajo peso y el sobrepeso pueden operar a través de mecanismos diferentes, y el modelo multinomial permite capturar estas diferencias específicas.

4.2. Ausencia de Multicolinealidad Severa (VIF)

- **Objetivo:** Verificar que no existan correlaciones excesivas entre las variables predictoras.
- **Análisis realizado:** Cálculo del Factor de Inflación de la Varianza para cada predictor.

```
##          GVIF Df GVIF^(1/(2*Df))
## edad_meses  1.020464  1      1.010180
## sexo        1.038027  1      1.018836
## peso_nacimiento 1.008600  1      1.004291
## edad_madre  1.035631  1      1.017660
## educacion_madre 1.040190  2      1.009900
## imc_madre    1.038492  1      1.019064
## zona         1.013440  1      1.006697
## lactancia_materna 1.020501  1      1.010198
## dietas_solidos 1.013418  1      1.006687
```

Todos los valores $GVIF^{1/(2 \cdot Df)}$ son < 2 , muy por debajo del umbral de 5, indicando ausencia de multicolinealidad problemática.

4.3. Verificación de la Linealidad en el Logit

- **Objetivo:** Evaluar si existe relación lineal entre las variables continuas y el logit de cada categoría.
- **Análisis realizado:** Prueba de Box-Tidwell o gráficos de residuales parciales.

H_0 : La relación entre el predictor continuo y el logit es lineal.

H_1 : La relación no es lineal.

```
##
## --- Evaluando variable: edad_meses ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 464.054597
## final value 464.006466
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.545546
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 463.479333
## final value 463.474077
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.545546
## converged
##
## --- Evaluando variable: peso_nacimiento ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 463.903844
## iter 20 value 463.566878
## final value 463.563911
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.560054
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 463.600798
## final value 463.592276
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.560054
## converged
##
## --- Evaluando variable: edad_madre ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 460.531906
```

```

## final value 460.450747
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 461.922125
## final value 461.919648
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 459.881825
## final value 459.869094
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 461.922125
## final value 461.919648
## converged
##
## --- Evaluando variable: imc_madre ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 464.483224
## iter 20 value 463.683615
## final value 463.671393
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 465.403773
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 460.694296
## final value 460.686589
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 465.403773
## converged
##
## --- Evaluando variable: ingreso_familiar ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 460.066679
## final value 460.066672
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 460.472620
## final value 460.472613
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 458.993891
## final value 458.985937

```

```

## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 460.472620
## final value 460.472613
## converged
##
## --- Evaluando variable: lactancia_materna ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 463.584406
## final value 463.583989
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 465.618886
## final value 465.615210
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 460.508559
## final value 460.504737
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 465.618886
## final value 465.615210
## converged

```

Método de evaluación: Se siguió el enfoque metodológico propuesto por Miranda (2024), que combina pruebas estadísticas formales con análisis comparativo de modelos. Para cada una de las seis variables continuas del estudio, se implementaron dos estrategias complementarias:

1. **Pruebas de razón de verosimilitud con términos cuadráticos:** Para cada predictor continuo, se comparó el modelo original con un modelo extendido que incluía un término cuadrático (χ^2) de la misma variable mediante pruebas de razón de verosimilitud (Likelihood Ratio Test).
2. **Análisis con splines cúbicos naturales:** Se ajustaron modelos alternativos utilizando splines cúbicos naturales (3 grados de libertad) para cada variable continua, comparándolos con los modelos lineales simples mediante pruebas estadísticas.

Ambos métodos permiten detectar desviaciones de la linealidad evaluando si la inclusión de términos no lineales mejora significativamente el ajuste del modelo. El criterio de decisión se basó en valores- $p < 0.05$ como evidencia de no-linealidad significativa.

##	Variable	P_cuadrático	P_spline	Conclusión
## 1	edad_meses	0.5833	0.7095	Linealidad aceptable
## 2	peso_nacimiento	0.3693	0.7476	Linealidad aceptable
## 3	edad_madre	0.2302	0.3925	Linealidad aceptable
## 4	imc_madre	0.1769	0.0511	Linealidad aceptable
## 5	ingreso_familiar	0.6663	0.5623	Linealidad aceptable
## 6	lactancia_materna	0.1312	0.0369	Posible no-linealidad

Interpretación de los resultados: Todos los valores-p obtenidos superan ampliamente el umbral de significancia estadística de 0.05. En particular:

- **Variables con valores-p muy altos (> 0.5):** `edad_meses` ($p = 0.71$), `peso_nacimiento` ($p = 0.75$) e `ingreso_familiar` ($p = 0.66$) muestran una fuerte evidencia a favor de la linealidad, con probabilidades muy bajas de que términos no lineales mejoren el modelo.
- **Variables con valores-p moderados:** `edad_madre` ($p = 0.23$) y `imc_madre` ($p = 0.18$ para términos cuadráticos; $p = 0.051$ para splines) presentan cierta tendencia hacia valores más bajos, pero aún dentro del rango de no significancia estadística.
- **Consistencia entre métodos:** Para las seis variables, ambos métodos (términos cuadráticos y splines) convergen en la misma conclusión, reforzando la validez de los resultados.

El caso del `imc_madre` merece atención especial: aunque el valor-p del spline (0.0511) se aproxima al umbral de 0.05, no alcanza significancia estadística. Esta proximidad podría sugerir una leve curvatura en la relación, pero insuficiente para justificar la inclusión de términos no lineales en el modelo final.

Conclusión: Con base en los resultados estadísticos presentados, **no se rechaza la hipótesis nula de linealidad** para ninguna de las variables continuas analizadas. La evidencia sugiere que las relaciones entre estos predictores y el logit del estado nutricional son aproximadamente lineales dentro del rango observado en los datos. Por lo tanto, el supuesto de linealidad en el logit se considera adecuadamente cumplido, validando la especificación funcional del modelo de regresión logística multinomial.

Limitación metodológica y consideración práctica: Cabe mencionar que la evaluación de linealidad en modelos multinomiales presenta desafíos inherentes, ya que debe verificarse para cada logit individual (cada categoría vs. la referencia). Los métodos aplicados aquí ofrecen una evaluación global razonable, aunque podrían no capturar no-linealidades específicas para transiciones particulares entre categorías. No obstante, la consistencia de resultados y los valores-p obtenidos proporcionan confianza en la adecuación del supuesto de linealidad para los fines inferenciales del presente estudio.

4.4. Independencia de Alternativas Irrelevantes (IIA)

- **Objetivo:** Evaluar el supuesto fundamental del modelo logit multinomial.
- **Análisis realizado:** Prueba de Hausman-McFadden.

Hipótesis

H_0 : El supuesto de Independencia de Alternativas Irrelevantes se cumple

H_1 : El supuesto de IIA no se cumple

Método: Se realizó la prueba de Hausman-McFadden comparando el modelo completo con un modelo reducido que excluye la categoría “sobrepeso”. Esta prueba evalúa si la exclusión de una alternativa afecta significativamente los coeficientes estimados para las alternativas restantes.

```
##
## Hausman-McFadden test
##
## data:  datos_mlogit
## chisq = -2.4882, df = 12, p-value = 1
## alternative hypothesis: IIA is rejected
```

Interpretación: No se rechaza la hipótesis nula ($p > 0.05$). No hay evidencia estadística contra el supuesto de IIA.

Conclusión: El supuesto de Independencia de Alternativas Irrelevantes se cumple. El modelo multinomial es apropiado para los datos.

5. Resultados del Modelo (Inferencia Estadística)

Aquí se presentan las estimaciones finales transformadas a valores interpretables (Odds Ratios).

5.1. Tabla de Razones de Riesgo Relativo (RRR / Odds Ratios)

- **Objetivo:** Presentar de forma clara y profesional los coeficientes del modelo con sus intervalos de confianza y significancia estadística.
- **Resultados a incluir:**
 - **Tabla 1:** Resultados del modelo de regresión logística multinomial, generada con `gtsummary`.
 - **Columnas:** Tres columnas principales:
 1. Variable predictora
 2. OR para “Normal” vs “Bajo Peso” (categoría de referencia) con IC95% y valor-p
 3. OR para “Sobrepeso” vs “Bajo Peso” con IC95% y valor-p
 - **Formato:** Variables significativas ($p < 0.05$) resaltadas en negrita o con asteriscos.
 - **Notas:** Incluir n de observaciones, pseudo R^2 , y pruebas de bondad de ajuste globales.

5.2. Interpretación de Predictores Significativos

- **Objetivo:** Explicar en términos sustantivos los hallazgos más relevantes del modelo.
- **Resultados a incluir:**
 - **Factores asociados a mayor probabilidad de estado NORMAL (vs bajo peso):**
 - * “La residencia en zona urbana aumenta 2.03 veces la probabilidad de tener peso normal frente a bajo peso (OR=2.03, IC95% 1.34-3.08, $p=0.001$)”
 - * “Cada año adicional de edad materna reduce levemente la probabilidad de peso normal (OR=0.96, IC95% 0.93-0.98, $p=0.015$)”
 - **Factores asociados al SOBREPESO (vs bajo peso):**
 - * “El peso al nacer muestra una asociación positiva fuerte: por cada kg adicional, aumenta 1.45 veces el riesgo relativo de sobrepeso (OR=1.45, IC95% 1.08-1.94, $p=0.013$)”
 - **Hallazgo inesperado:**
 - * “Contrario a lo esperado, la educación superior materna se asocia con mayor riesgo relativo de bajo peso (OR=0.21, IC95% 0.06-0.78, $p=0.020$) cuando se compara con el grupo de sobrepeso”

6. Visualización de Efectos Marginales y Probabilidades Predichas

Se traduce la estadística compleja a gráficos intuitivos que facilitan la interpretación sustantiva.

6.1. Dinámica del Ingreso Familiar y Zona Geográfica

- **Objetivo:** Mostrar cómo interactúan el nivel socioeconómico y el contexto geográfico.
- **Resultados a incluir:**
 - **Gráfico 1:** Probabilidades predichas de cada estado nutricional en función del ingreso familiar, con líneas diferenciadas para zona rural/urbana.
 - **Interpretación:** “En zona rural, la probabilidad de bajo peso se mantiene alta (>40%) incluso con ingresos medios, mientras que en zona urbana disminuye rápidamente por debajo del 20%”

6.2. Impacto del Nivel Educativo Materno y Perfiles de Riesgo

- **Objetivo:** Visualizar diferencias en el perfil de riesgo nutricional según características clave.
- **Resultados a incluir:**
 - **Gráfico 2:** Diagrama de barras mostrando probabilidades predichas por nivel educativo materno.
 - **Gráfico 3:** Gráfico de perfiles (spaghetti plot) mostrando cómo cambia la probabilidad de bajo peso según edad materna e IMC materno.

7. Evaluación del Ajuste y Capacidad Predictiva

7.1. Bondad de Ajuste Global

- **Resultados a incluir:**
 - **Pseudo R²:** Reportar McFadden, Cox-Snell y Nagelkerke.
 - “El Pseudo R² de McFadden (0.08) indica un ajuste moderado, típico en modelos de ciencias sociales/salud”
 - **Test de razón de verosimilitud:** $\chi^2=50.67$, $gl=22$, $p<0.001$ (ya calculado)

7.2. Capacidad Predictiva y Validación

- **Resultados a incluir:**
 - **Tabla 2:** Matriz de confusión con clasificación real vs. predicha.
 - **Métricas de rendimiento:**
 - * Exactitud global (Accuracy): XX%
 - * Sensibilidad por clase: Bajo peso (XX%), Normal (XX%), Sobrepeso (XX%)
 - * Especificidad por clase
 - * Estadístico Kappa de Cohen (acuerdo más allá del azar)
 - **Validación cruzada:** Resultados de validación con k-fold (ej. 10-fold) para estimar el error de generalización.

8. Discusión, Conclusiones y Aplicaciones

8.1. Síntesis de Hallazgos Principales

- Resumen en 3-4 puntos clave:
 1. “Se confirma la doble carga de malnutrición: 30% bajo peso y 12% sobrepeso”
 2. “La zona rural y menor ingreso son factores de riesgo consistentes para bajo peso”
 3. “El peso al nacer emerge como predictor importante de sobrepeso”

8.2. Discusión en Contexto Teórico y Práctico

- Comparación con literatura previa
- Explicación de hallazgos inesperados (ej. educación materna)
- Mecanismos potenciales que explican las asociaciones encontradas

8.3. Limitaciones del Estudio

- Datos simulados (si aplica) y limitaciones de validez externa
- Variables no medidas (ej. seguridad alimentaria, calidad de dieta)
- Posible causalidad reversa en algunas asociaciones

8.4. Implicancias para Políticas Públicas

- Recomendaciones específicas y accionables:
 - “Priorizar intervenciones nutricionales en zonas rurales con enfoque en ingresos”
 - “Programas de seguimiento para niños con alto peso al nacer”
 - “Estrategias diferenciadas según nivel educativo materno”

8.5. Caso Práctico: Predicción para Perfiles Específicos

- Ejemplo 1: “Niño rural, ingreso bajo, madre con educación primaria”
 - Probabilidades predichas: Bajo peso (65%), Normal (30%), Sobrepeso (5%)
- Ejemplo 2: “Niño urbano, ingreso alto, madre con educación superior”
 - Probabilidades predichas: Bajo peso (10%), Normal (85%), Sobrepeso (5%)

8.6. Recomendaciones para Investigación Futura

- Sugerencias metodológicas: estudios longitudinales, inclusión de variables adicionales
- Sugerencias sustantivas: análisis por subgrupos, estudios cualitativos complementarios

3. Regresión Logística Ordinal con Odds